BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e26168
doi: 10.3897/biss.2.26168

OPEN ACCESS

Conference Abstract

Discussion of the Method for Constructing Animal Traits

Jiangning Wang‡, Congtian Lin‡, Yan Han‡, Liqiang Ji‡

‡ Institute of Zoology, Chinese Academy of Sciences, Beijing, China

Corresponding author: Jiangning Wang (wangjn@ioz.ac.cn), Liqiang Ji (ji@ioz.ac.cn)

Received: 25 Apr 2018 | Published: 25 Apr 2018

Citation: Wang J, Lin C, Han Y, Ji L (2018) Discussion of the Method for Constructing Animal Traits. Biodiversity Information Science and Standards 2: e26168. https://doi.org/10.3897/biss.2.26168

Abstract

Trait data in biology can be extracted from text and structured for reuse within and across taxa. For example, body length is one trait applicable to many species and "body length is about 170 cm" is one trait data point for the human species. Trait data can be used in more detailed analyses to describe species evolution and development processes, so it has begun to be valued by more than taxonomists. The EOL (Encyclopedia of Life) TraitBank provides an example of a trait database.

Current trait databases are in their infancy. Most are based on morphological data such as shape, color, structural and sexual characteristics. In fact, some data such as behavioral and biological characteristics may be similarly included in trait databases.

To build a trait database we constructed a list of controlled vocabulary to record the states of various terms. These terms may exhibit common characteristics:

1. They can be grouped as conceptual (subject) and descriptive (delimiter) terms. For example, in “the shoulder height is 65–70 cm”, "shoulder height" is the conceptual term and "65–70 cm" is the descriptive term.
2. Conceptual terms may be part of an interdependent hierarchical structure. Examples in morphology, physiology and conservation or protection status, demonstrate how parts or systems may be broken into smaller measurable (quantifiable) or enumerable pieces.

© Wang J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

3. Descriptive terms will modify or delimit parameters of conceptual terms. These may be numerical with distinguishing units, counts, or other adjectives or enumerable with special nouns.

Although controlled vocabularies about animals are complex, they can be normalized using RDF (Resource Description Framework) and OWL (web ontology language) standards.

Next, we extract traits from two main types of existing descriptions.

1. tabular data, which is more easily digested by machine, and
2. descriptive text, which is complex.

Pure text often needs to be extracted manually or by NLP (computerized natural language processing). Sometimes machine learning methods can be used. Moreover, different human languages may demand different extraction methods.

Because the number of recordable traits exceeds current collection records, the database structure should be optimized for retrieval speed. For this reason, key-value databases are more suitable for storage of traits data than relational databases. EOL used the database Virtuoso for Traitbank, which is a non-relational database.

Using existing mature tools and standards of ontology, we can construct a preliminary work-flow for animal trait data, but some tools and specifications for data analysis and use need to await additional data accumulation.

Keywords

animal, trait data, key-value database, ontology

Presenting author

Jiangning Wang